EMMC-CSA

European Materials Modelling Council

Report on Workshop on Interoperability in Materials Modelling

Cambridge, 7/8 November 2017

TABLE OF CONTENT

The EMMC-CSA project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 723867

1. Executive summary

The workshop was attended by 66 experts representing a wide range of stakeholder communities, including academic and commercial materials modelling software owners (CULGI, SCM, Micress, 3DS, Siemens PLM etc), manufacturing industry (Dow, Johnson Matthey), modellers covering different types of models and applications, repository owners (Granta Design, NOMAD project), academic and commercial data and science/informatics software owners/consultants (Enthought, Data Stories, Intellegens, Osthus). It also brought together a number of current EU projects, the three DGCNECT COEs in the materials modelling field (NOMAD, MAX, E-CAM) and EUDAT (including a presentation).

The objective of the workshop was for EMMC to seek the support of and endorsement by the wider materials modelling community for the European Materials Modelling Ontology (EMMO). EMMC is also gathering requirements and outlining plans for interoperability between data repositories and marketplaces.

The workshop discussed recent developments in interoperability approaches in materials modelling and related fields, following on from discussions at the First EMMC International Workshop in April 2017 Vienna. In particular, this workshop focused on semantic interoperability based on a European Materials Modelling Ontology (EMMO) which is currently being developed. EMMO will form the basis for interoperability and domain specific metadata.

Documentation, repositories and databases were also discussed, including initial requirements for cataloguing simulations in data repositories via a MODA, and the support for their elaboration by a portal.

The major outcome of the workshop was the wide agreement on the **need for ontologies in materials modelling**. Moreover, there was a call for an integrated effort to develop **ontologies for the whole field**, including materials characterisation and modelling all the way to chemicals and materials development in industry. Such a development is regarded as key to success in digitalisation. The draft **EMMO was well received** and its current development was widely endorsed. The **status**, **requirements**, **expectations**, **benefits**, as well as **potential pitfalls** of ontologies were **discussed in detail.** This revealed that there was a need to establish **use cases for EMMO** together with industry. These will serve to establish more detailed requirements and demonstrating the capabilities of EMMO to all stakeholders. Also, there was the request for establishing **governance for EMMO** with clear procedures for contribution, communication and other types of interaction, which was deemed important by all participants. The need for a **common semantic basis** and hence ontologies that drive the marketplace platform that will link all stakeholders (providers and buyers) and integrate translation and decision support was highlighted. There is good evidence that the **Use of the RoMM terminology** during communication about modelling and simulation is **spreading well** beyond EU projects. Finally, there is good evidence that the Documentation of simulations **using** MODA is now widely **endorsed** and is supporting communication and the development of the MODA portal

Workshop Webpage:

https://emmc.info/emmc-workshop-on-interoperability-in-materials-modelling/
The webpage includes the programme and links to the presentations.

2. Report

2.1 Background: Taxonomy and Ontology

Taxonomies contain the definition of keywords in a domain and a classification. Ontologies specify the relations. Taxonomies are at the heart of an ontology and even often considered to be part thereof; Ontologies are vital for facilitating communication and knowledge exchange between communities as they enable mapping between different taxonomies.

2.2 Status and requirements for interoperability

2.2.1 Manufacturing industry

In order to accelerate innovation, industry requires data and resources relevant for the development of new and enhanced materials and new products to be readily available and deployable. Industry needs to be agile and fast in its response to customer and market needs, which translates into the requirement to be able to decide upon and utilise appropriate models quickly but also to link modelling and experimental data. It means drawing on existing knowledge and avoiding unnecessary repetition in R&D effort. In order to achieve that and **accelerate the materials and product development process**, industry needs to make information discoverable and curated and provide annotated data with provenance information.

Manufacturing Industry actively seeks the ability to perform rapid inference exercises. These are the process needed for reaching a conclusion on the basis of evidence and reasoning, e.g. answering the question: "A customer wants certain materials' key performance attributes: what application gets them there?". Manufacturing industry requires systematic recommenders, which are systems that produce a list of recommendations when asked for performance requirements. Both these trends must be embraced and established.

Today, however, most **data and expertise are scattered** in various communities and geographies; there are severe problems including ease of access to data and insufficient annotation and context of data which hampers the possibility to fully leverage the inherent benefits of such hidden knowledge [see e.g., D3.8].

The manufacturers attending this meeting clearly stated that ontologies provide a good way towards guaranteeing seamless interoperability. However, there is very little that materials manufacturing industry can rely on in terms of knowledge and information systems such as ontologies in public domain knowledge. Hence, manufacturing companies currently resort to creating **domain specific, internal (private) ontologies** that require significant effort and hence are usually limited to large enterprises. An external, community endorsed ontology should provide the backbone and be able to host such a domain specific ontology. There are no public ontologies available to simply enable industry to plug-in their domain specific or internal ontology and to create an interoperability basis.

In the question and answer sessions after each of the manufacturing talks, it emerged that to optimise database use, ontologies need to be involved already when data are generated. The industrial stakeholders acknowledged that the MODA is one such example of structuring and annotating data in order to make them fit for interoperability. Open Simulation Platforms built on the basis of ontologies are an ambition to provide tools that allow the exchange of knowledge and data between modelling tools, data repositories and communities.

Ontologies also have an impact on the emerging field of materials informatics supporting the use and integration of data-based modelling (including data-mining, machine learning, etc.) in materials and product development. This has been proven especially in Life and Chemical Sciences where such ontologies facilitate the exchange of knowledge and discovery of new pharmaceuticals.

The manufacturers were pointing out that an ontology should support inference and recommenders, in fact, it is a “must-have”. **Ontologies should help with spanning a bridge to data and knowledge** in the public domain and thus embrace what is already established.

To fill this existing gap, the **development of the EMMO was welcomed** by all industrial stakeholders present at this workshop.

2.2.2 Software Owners

Software Owners need to serve their customer’s needs for improved interoperability, integration and annotation of data. Industrial software development in the field of materials modelling is driven by customers demanding more multiscale modelling workflows preferably with as little user interference as possible. Furthermore, addressing manufacturing problems requires the utilisation of **multiple models in linked and coupled workflows**. These workflows require in turn the exchange of information and data between models and hence interoperability is essential. Software Owners need to achieve this in a **way that** improves the long-term health of their software**. Creating single point solutions and one-off interfaces between codes is clearly not sustainable.**

2.2.3 State of the Art

Regarding documentation of simulations using physics-based or data-based materials models, the state of the art is the MODA. MODA provides a framework for consistent documentation, ensuring that all aspects of a simulation are covered. It is based on the terminology and classification of materials modelling outlined in the RoMM. Several participants of this workshop commented on the usefulness of MODA and its graphical representation. People often criticised the amount of work involved with filling in MODA tables, hence an online portal will be established to ease the process and digitalise the data. The **MODA portal** application will provide also a prototypical example of a web based app that allows the entire community to use an ontology based tool without having to know the details. It will enable users to conduct searches, compare data, and find information in a semantic way.

Considering materials modelling data repositories, the NOMAD project gave a good insight into how to **store data in a syntactic context** and make them searchable by using identifiers. A **user-friendly**

API is key and so is interactive online documentation about the metadata. NOMAD developed a system of metadata, in particular for the solvers of electronic and atomistic models. Annotating material structures (e.g. crystal structures) not only with computed properties but also automatically with full provenance information of the simulations was discussed in the presentation by Marzari.

Materials modelling interoperability largely happens via **file based exchange** involving certain formats and at best some partial metadata, that is not always properly documented as it is not guided by an ontology. The N^2 problem of writing file based interfaces or wrappers between every two pieces of code was highlighted and it is understood that this could be overcome by a common reference, established on the basis of an ontology.

At present, many software vendors offer solutions to make their software interoperable with 3rd party products (customer-driven), but these are not built on ontology.

2.2.4 The way forward: Ontologies

Vendors recognize that an **ontology** would make building and maintaining interoperability more **sustainable**. Hence, ontologies are a topic of much interest in particular to larger software owners, including Dassault Systèmes and Siemens PLC, as evidenced in their workshop discussion contributions.

Developing ontologies requires substantially more effort than developing taxonomies. The **value of taxonomies** is that a **common language** is established and communication can happen on that basis, both by humans as well as machine processing. However, **ontologies elaborate** this by **agreeing on the relations** between universals. Therefore, **ontologies enable interoperability** between codes and models which can be handled via a common reference, i.e. guaranteed by common standards and related interface wrappers built on these standards.

Ontologies allow annotating (“decorating”) data (numbers, representations of things, such as models, boundary conditions, materials and their properties) in a univocal manner allowing their interpretation both by humans and most importantly by machines. The latter means that code can be written to allow autonomous or at least semi-autonomous decision making leading to inference and recommendations on how to utilise the data.

Ontologies provide software owners and developers with a high-level description of the domain knowledge and how the various parts of knowledge are interconnected. This allows the generation of cleaner and more sustainable and interoperable codes.

Nevertheless, it is still a field of research and development and details about the actual impact and value in the materials space have yet to emerge. Some examples of ontology based projects were discussed in the workshop. These included project specific ontology developments that utilise the structured approach to provide interoperability and reduce errors e.g. in unit conversions (see presentation by Maciol). The Allotrope project aims to provide data interoperability for analytical laboratories. Driven by specific use cases and requirements, the first solution built was based on a taxonomy rather than a full ontology. However, it became clear later that this was insufficient and a full ontology was developed.

Industry and SWO request ontologies as they wish to do operations using the relations (knowledge engineering), for example in Business Decision Support Systems (BDSS).

The nature of the relations significantly influences the type of ontology and hence the actual organisation of knowledge based on such an ontology.

Automatic code generation can be used to lower costs, guarantee better software development standards and accelerate the development of interoperable interface software wrappers as is demonstrated, for example, in the EU projects SimPhoNy and NanoSim. Also, the outputs/results from a simulation or lab experiment can be annotated automatically.

A **large range of tools** can be built that **rely on data organised via an ontology**. These include:

a) Query tools

b) Reasoners

c) Comparison tools

e) Validators (of the data)

f) Visualization/navigation tools

Furthermore, ontologies can support **cross-domain interoperability**, i.e. the exchange of data across domains and communities, each with its own domain ontology. If the domain ontologies share an upper ontology, the mapping between data from domains could be done at least semi-automatically (see also section 7.8).

2.2.5 Requirements and expectations for the ontology

Building an ontology on an upper level is important to arrive at an ontology that can be linked to other domain ontologies. BFO is such an upper level ontology. EMMO is BFO-based, and so are chemistry ontologies (e.g. CHEBI), recent efforts in Materials Ontology development (e.g. MatOnto) and company internal efforts (e.g. at Dow). EMMO should capture the highest levels of the materials and materials modelling field. In a sense, EMMO is an intermediate between BFO (as upper level ontology) and specific ontologies in the areas of modelling and materials science. Building specific, detailed ontologies (public or company internal) on this basis will automatically lead to a degree of interoperability.

It would be of interest, that the EMMO becomes BFO-conformant as some existing ontologies are in the process of being re-engineered to be BFO-conformant such as SLACKS (Laminated Composites, UMass) and FGMO (Functionally Graded Materials NCOR, Milan Polytechnic). CHEBI (Polymers, EBI) is already BFO-conformant. The EMMO may also look into a meta- and a detailed ontology called Multiscale Modelling Formal Ontology MMFO, and OntoCAPE (general purpose ontology for applications domain of Computer Aided Engineering). Being BFO compliant puts EMMO into a strong position to leverage increased support and adaptation that even goes beyond the EU, and as such increases European leadership internationally.

The European Materials Modelling Ontology (EMMO) is based on BFO and captures the description of systems and materials to be modelled, taking different levels of granularity of description into account. Modelling is represented consistently with the RoMM and the CWA of materials modelling terminology, classification and metadata. **EMMO** is work in progress but the structure so far presented was **widely endorsed** by the workshop participants. Further developments are eagerly awaited and **governance will need to be established** to manage participation of stakeholders. Similar to the life science ontologies, one may want to go via an **editorial board**, so that the EMMO can be entered to the BFO.

A key factor is providing interoperability between data repositories, where EMMO plays a central role. This part of work is extensive, due to the amount and diversity of available repositories. Similar to interoperability between simulation tools, one requires interoperability between the access schema for databases, and this can be achieved by a common reference API. Such an API can be also standardised and open so that each repository can implement it and thus, allowing the interlinking of repositories. This will enable the creation of a **critical mass of knowledge and information** allowing the rapid development of a new kind of tools based on machine learning and big data analysis to discover new materials and products faster. Such approaches are at the centre of various EU projects (FORCE, MARKETPLACE, VIMMP) but more interaction with the community and especially commercial database providers is needed.

The EMMO should be tested to ensure that it covers all statements that may occur within Materials Modelling. However, the **ontology** should **not be too complex**, it should be kept to the minimum that is needed for the community. This means, the EMMO will comprise a minimum set of “words” and then relations should lead to all the other “words”.

End users may not want to become proficient with ontologies. They rather prefer to use it similar to **a black** box and get the benefits it offers without having to deal with the underlying framework. An ontology basis is for computers. It allows storing and managing knowledge and it transforms raw data into knowledge based entities that can be used and reused in different contexts. Integration of diverse data and interoperability are the result of providing meaning and context of data, and thus, transforming it into knowledge. The ontology needs to be strong on the definition of rules to **support reasoning, recommendations and decisions** (see also section 2.1.1). However, for complex systems or workflows the reasoner may require a lot of computer power and time which will have to be considered in the ontology design. Furthermore, new logic and algorithms need to be developed to utilise the semantic power to its full extent; this requires a paradigm shift in the way materials modelling tools are developed. Ontology driven Open Simulation Platforms can become key to providing such tools and new paradigms.

A semantic framework can aid in the post-processing of data without errors (e.g. concerning units), if the relevant concepts are in the ontology. However, without the relevant and correct ontology, the computer cannot understand that there are dependencies between properties, and complex interpretations may be out of scope. It was also highlighted that postprocessing of data is not merely a numerical exercise but it does entail a computational representation of the science of coupling and linking, and as such ontologies for post-processors should be included in the EMMO.

2.2.6 How to make the EMMO a success?

The Allotrope Project (also BFO based) is an example of how to establish an ontology driven by the community. The Allotrope project was created by pharma companies since their labs, using equipment from different vendors, produced non-interoperable data. It is strongly driven by an end-user community from the pharmaceutical industry and supported by instrument manufacturers. The ontology building has been driven by specific use cases and has been in the hands of a dedicated team.

Experience was also shared by Barry Smith regarding the Industry Ontology Foundry (IOF) initiative (NIST, Airforce Research Lab, industry and academia) which created a suite of interoperable, high quality ontologies covering the domain of industrial (especially manufacturing) engineering. From this effort, one can deduct, that they should be small but **easily extensible** for different companies, products, etc.

Large materials manufacturing companies, as stated above, invest some internal resources to create a domain specific materials ontology. However, in contrast to the pharmaceuticals industry (see Allotrope.org) they do not seek collaboration with other companies in how to achieve this. Broadly speaking, materials ontology is **not yet a main talking point** in industry.

The **EMMC-CSA needs to involve the users** as SWOs will only respond if there is evidence of a high demand for interoperability supported by ontologies. The SWOs were adamant that they will in any case not abandon their past investments. They require therefore **interface wrappers** to be developed that translate outputs of one tool to inputs for another tool and thus span the bridge between different models. Such software interface wrappers render existing and legacy code interoperable and hence allow the vendors to still use their own data formats internally. As wrappers may provide an added channel for interoperability at low cost and efforts, the EMMC should also facilitate them.

EMMO should combine use cases with the need to **establish broad concepts** that are applicable to the whole field. Thus, an important step will be for the EMMC-CSA to **identify case studies** that demonstrate within an organisation, how domains can be opened up to reach interoperability and demonstrate the business benefits. Such case studies can then be used to demonstrate the usefulness of the EMMO to a wider audience and make it a success. Manufacturing industry indicated their support for this course of action.

2.2.7 Integration and interoperability in Materials Modelling Marketplaces

Materials modelling is moving from local infrastructures to the **worldwide community** as evidenced by different projects and companies, who shared their experiences during the workshop. Providing a single/unified marketplace experience for all of materials modelling is the challenge of the years to come and two new projects (MARKETPLACE, VIMMP) will start in 2018. Ontology development will be key to achieving powerful and scalable marketplaces that enable to plug-in a wide range of codes and models, build workflows easily, provide knowledge-based translation capabilities, powerful data mining and analytics and results that the user can trust. EMMO will be an important backbone in this process.

3. Conclusions

In addition to the Major Outcomes already outlined, the following actionable conclusions are drawn:

- Work on EMMO should be prioritised to get a version EMMO 1.0 out in the open, establishing the high level of an ontology.
- Guidance for the development of EMMO, demonstration of applications and value should be based on use cases. Invite manufacturing industry for the definition of the cases.
- The EMMO governance structure needs to be clarified. One or only a few persons should oversee the top level, and branches may be covered by domain experts. Mechanisms for contributions and involvement of potential users need to be put in place.
- It needs to be clarified how different types of stakeholders interact with EMMO, including software owners. How can EMMO integrate with existing tools and software infrastructures?
- The EMMO project needs to work closely with EU marketplace projects.

4. Annex: Workshop Discussion Notes

4.1 Topic of the workshop

This joint workshop of the EMMC "Interoperability and Integration" and "Repositories and Marketplaces" Working Groups focuses on interoperability aspects covering materials models, repositories and marketplaces.

4.2 Introduction to interoperability

All stakeholders of materials modelling face barriers regarding access to and utilisation of the wide range of models, data and software tools, as well as materials modelling information and expertise. More efficient access and effective utilisation requires interoperability between all of these parts.

Interoperability brings materials modelling communities closer together, boosts collaborative science and enables increased leverage of advances in materials modelling for the benefit of European Industry.

Interoperability is also very closely connected to digitalisation. Progress from electronic formats to truly digitalised representations enables new ways of operating, collaborating and benefitting from materials modelling.

Digital Materials Modelling Marketplaces are an example of gathering all parts and allowing them to cooperate.

4.3 Materials Modelling Marketplaces

Materials Modelling Marketplaces are digitalised systems which integrate a range of tangible and intangible components to support innovation based on materials modelling. Marketplaces utilise web-based platforms in order to link various materials modelling activities including repositories, modelling workflows, simulation tools, expertise, training, translators, etc.

Existing and emerging repositories and materials modelling marketplaces call for additional actions ensuring coherency and efficiency of information management and exchange. In particular linking various marketplaces and data repositories requires interoperability to facilitate common unified access and retrieval of data and information.

There is also a need for efficient and lean management and curation of data and knowledge across different platforms. These, in turn, pose additional requirements for deep interoperability that go beyond models, reaching out into data and information management in general.

4.4 Different types of interoperability

In general terms, interoperability is defined as the ability to make data collected in one system usable within the framework of a second system without further intervention by humans.

In our context, we define interoperability as the ability of a model (either physics or data based model) to make use of information generated by a simulation with another model or from a database (see Review of Materials Modelling Version 6).

While interoperability refers in general only to the *exchange of information* without reference to *how this is implemented*, in practice, there are *three main levels of interoperability realisations*, **syntactic**, **semantic**, and **cross domain** as shown schematically in Figure 1.

Syntactic interoperability means that you make your tools understand the input and output from each other. This works, but only case-by-case and is not robust to changes.

Semantic interoperability is the ability to communicate shared meaning. It provides much more flexibility but requires means to describe the data (using metadata) that link each data element to at least a controlled shared vocabulary, and preferably an associated ontology. Typically, these vocabularies and ontologies are limited to a certain domain. Semantic interoperability enables codes to be written that can operate on various data resources and add "understanding" (semantics!) to the data, promoting them to the level of *information*.

Cross-domain interoperability refers to an even higher level of semantics and enables interoperability between domains and ontologies.

The objective of this meeting

The objective of this meeting is to discuss how semantic and cross-domain interoperability could be addressed by EMMC. First steps are already undertaken and the European Materials Modelling Ontology (EMMO) and the basic metadata schema (described below) are two ongoing efforts to address semantic and cross-domain interoperability.

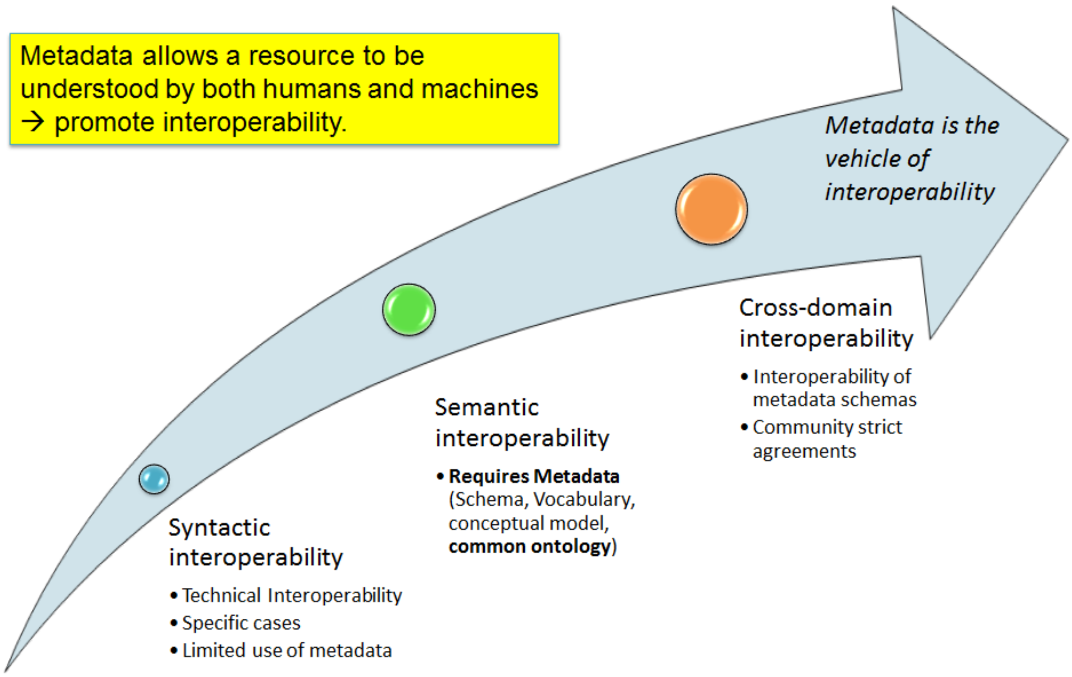

Figure 1. Levels of interoperability.

4.5 European Materials Modelling Ontology (EMMO)

EMMO is an emerging action to bring to materials modelling the same benefits that similar ontologies have brought to the fields of bio-informatics and cheminformatics.

EMMO is an ongoing effort to create an ontology for materials modelling and thereby pave the road for semantic interoperability within the field of materials modelling. The aim of EMMO is to be generic and to provide a common ground for describing materials models and data that can be adapted by all domains within EMMC.

As illustrated in Figure 2, there are several levels of structuring before reaching a proper ontology. EMMO builds further on the vocabularies and taxonomies already established in the RoMM[i] and further elaboration by the EMMC through the CEN Workshop on materials modelling terminology, classification and metadata[ii].

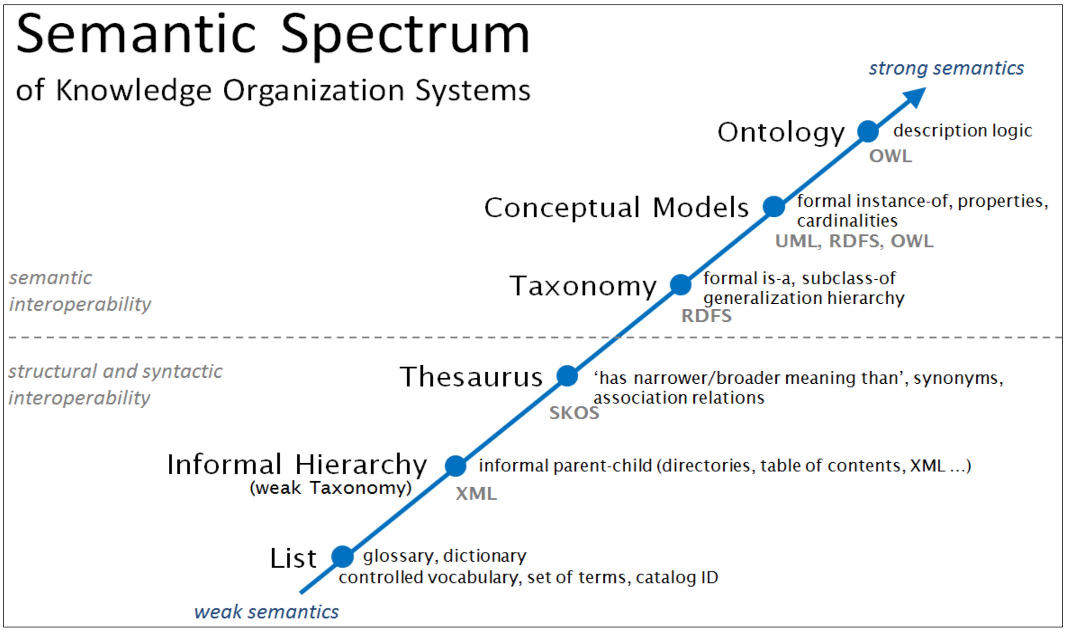

Figure 2. *Levels of semantic systems. Source: Geoff Gross, Osthus.*

The EMMO is built on top of the Basic Formal Ontology (BFO)[iii]. BFO is a top-level ontology which includes the highly general representation of categories and relations common to all domains, and is built in particular for the 'material world'. Its widespread use will ensure that domain specific ontologies are built in a consistent manner and share common high-level universals.

BFO will be introduced as part of the presentation by Barry Smith.

See also http://ontology.buffalo.edu/smith/ for a wide range of resources.

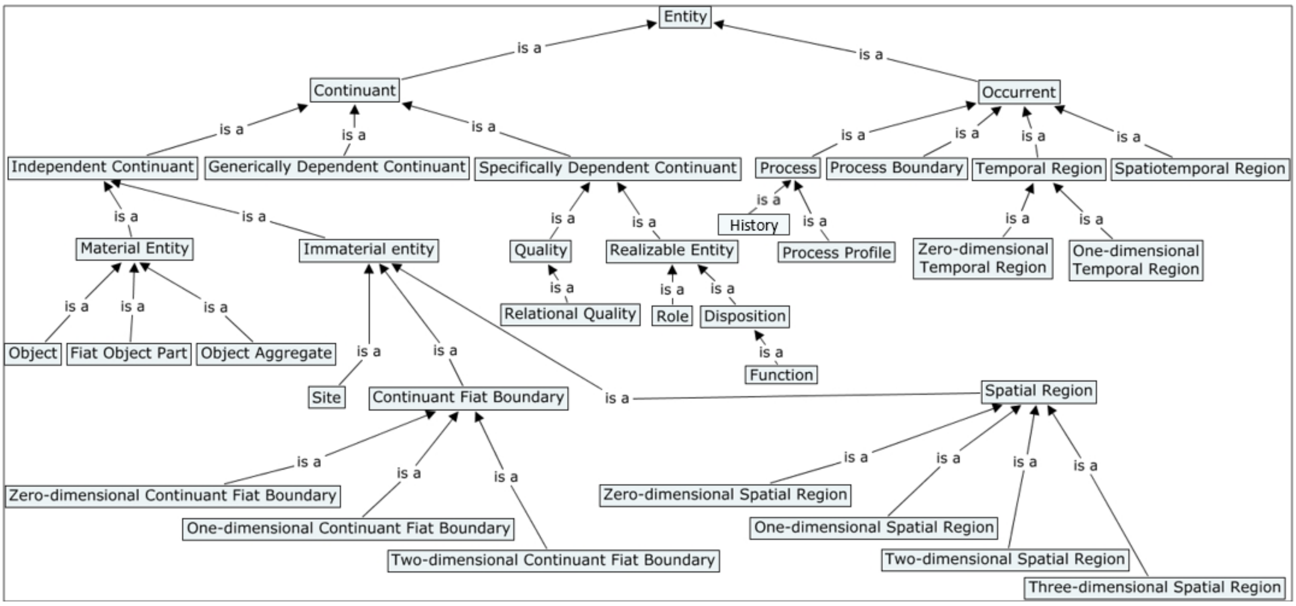

Figure 3. Taxonomy for the Basic Formal Ontology (BFO). Source: BFO Reference[iv]

There are a wide range of ontologies based on BFO, in particular Gene Ontologies, Chemistry Ontologies etc. However, there are also ontologies of importance to our field that are not BFO based, such as OntoCAPE (chemical engineering/processing) and Engineering Materials Ontology (see: CEN CWA 16762:2014).

The top-level universals of BFO are Continuants (things that exist throughout time) and Occurrents (Processes in time) (see Figure 3). EMMO development is concerned at this stage with Continuants, which include branches covering the Material Entities that we want to model, and the types of Objects used to describe them, the physical laws underlying the physics based models, the models (physics and data based) and the mathematical representations.

Figure 4 shows the branch of the draft EMMO detailing the taxonomy for the Material Entity branch.

The 'engineered material' represents the 'User Case Material' in the MODA that is to be studied. An instance of the 'engineered material' can be a particular sample of alloys, metals, plastics, etc. When describing and modelling such a material, granularity and perspectivism are important. The EMMO includes the objects representing four granularity levels: subatomic, atomistic, mesoscopic and continuum. The modeller hence describes the 'engineered material' by means of these types of objects and/or their respective aggregates. For example, a piece of metal can be described as a solid continuum object or an atom aggregate. Not included in the view in Figure 4 are the parthood relationships. Objects can have other objects as parts. For example, a molecule object has atoms as parts.

Further development of granularity and perspectivism are issues high on the agenda to be incorporated. Also, the ontology of the knowledge issue "model" (a Generically Dependent Continuant) is on the agenda.

The current status of EMMO will be presented at the workshop.

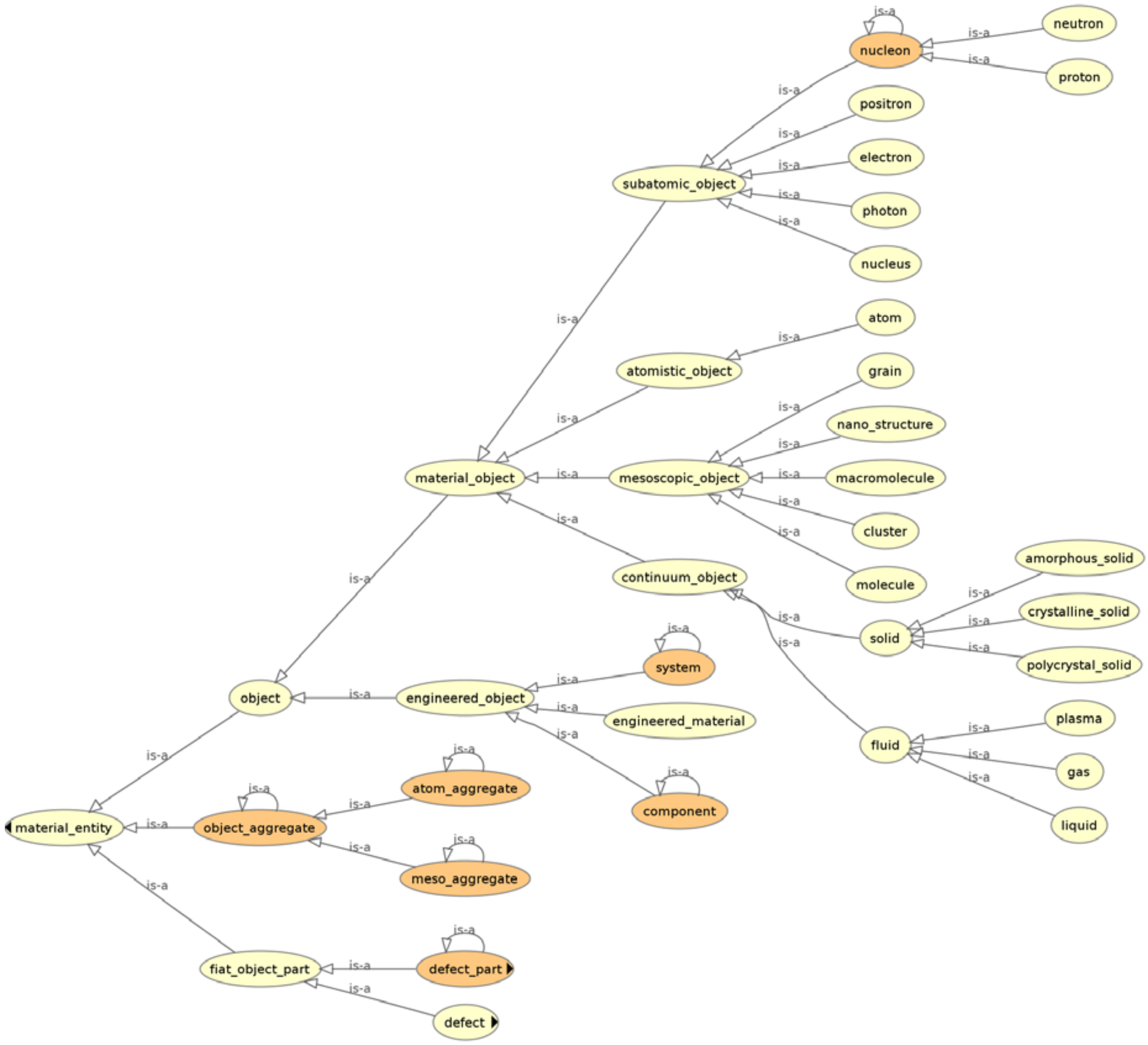

Figure 4. Branch of EMMO, showing the taxonomy for the Material Entity branch. This is "work in progress".

4.6 Objectives of EMMO

The goal is to demonstrate the efficiency and practicality of an ontology based approach to interoperability.

- Simplified and more efficient communication across the field of materials modelling.
- Improved communication of materials modelling across other science and engineering disciplines.
- Improved communication of materials modelling to industry.
- Easier use and re-use of data from, and information about materials modelling.
- Interoperability, including between models and databases and marketplaces.
- Creating a semantic reference for the specifications of file-based schema ("data formats") such as the hierarchy and keywords in HDF5.
- Support the process of translating industrial problems into problems that can be simulated with materials models.

- Assist workflow development where several models can interoperate, including on marketplaces.
- Streamline data curation across multiple repositories with provenance and data governance models.

4.7 Discussion points and questions

1. **Ontology has had a big impact on biosciences over the last decade allowing exchange and curation of data and enabling rapid innovation.**
    - How can materials sciences adopt a similar approach?
    - What are the already ongoing efforts in engineering and materials?
    - What are the low hanging fruit?
2. **EMMO is designed to provide the semantic framework that software vendors can use to build interoperable solutions.**
    - What are the software owner needs?
    - What collaborations or consortia can be put in place to further the common goals?
    - What can we learn from other fields and initiatives (e.g. Allotrope)?
3. **EMMO should support industry in managing and utilising models and data better as 'digital assets'.**
    - What are industry requirements?
    - How can collaborations with industry built to ensure this goal is achieved?
4. **EMMO should facilitate the integration of materials modelling with other digitalisation efforts in industry.**
    - What efforts are needed to make this happen in an efficient and effective manner?
5. **EMMO is a combined top-bottom and bottom-up activity, it requires commitment from the whole community to endorse it and use it.**
    - What are the potential bottlenecks and barriers for a wide endorsement and use?
    - What governance structure should be put in place of the EMMO?
6. **Cross domain or semantic interoperability requires ontology, but this is usually used to annotate data, while data repositories and marketplace need to cater for the curation and transfer of data.**
    - Is this sufficient to achieve interoperability and linking between repositories?
    - Do we need an ontology of "operations, methods, and actions" on data, in terms of post, pre-processing, and accessibility? (these can be manifested, e.g., by ontology of application programming interfaces (API's))
7. **Marketplaces contain much more than data: also, models, workflows and "social networking" components:**
    - Does the interoperability between marketplaces require an auxiliary ontology for workflows? Should this be part of the EMMO?
    - Are there ontologies out there already for workflows that we can use?
    - How to create and curate the mathematical representations of the model equations?

4.8 Future outlook – basic metadata schema

The EMMO aims to be an ontology that covers the domain of materials modelling, and in real applications there will be a need to connect to other domains. In order to implement cross-domain interoperability we must be able to translate between ontologies, which requires a common language for describing each ontology. The Web Ontology Language (OWL) is one such a language, which is used by Protégé to express BFO. OWL itself uses the Unified Modelling Language (UML) to define its structural elements, which is based on the Meta-Object Facility (MOF). This is a very general framework that is difficult to grasp for people not specialised in computer science. We therefore suggest creating a simplified subset of MOF, called **basic metadata schema**, with the following three structural elements:

- keyword-value pairs (properties)
- dimensions (provides arrays and lists)
- relations

In addition to means of uniquely referring to the metadata descriptions (as name, version, namespace triplets) and sane versioning of changes. We believe that these structural elements are easy to grasp and yet sufficient to describe all relevant metadata, ontologies and actual (realised) data. A formal UML diagram for the proposed basic metadata schema is shown in Figure 5.

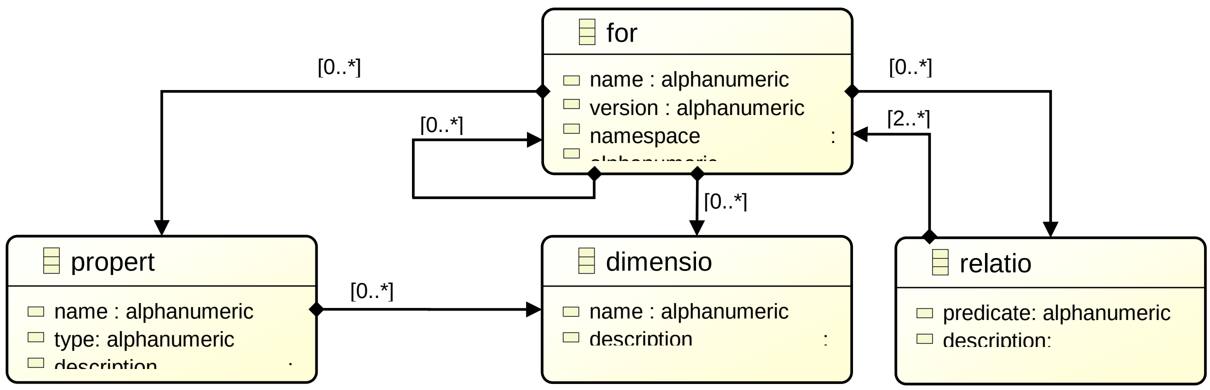

Figure 5: A formal UML diagram for the proposed basic metadata schema.

4.9 Appendix: Terminology

Metadata can be defined as "data describing the context, content and structure of records and their management through time"[v]. They provide information that allows for categorization, classification and structuring of data[vi]. A well-established example is the Crystallographic Information File (CIF)[vii] which provides metadata for atomistic structures and properties.

A **metadata schema** can be defined as "a logical plan showing the relationships between metadata elements, normally through establishing rules for the use and management of metadata specifically as regards the semantics, the syntax and the optionality (obligation level) of values"[viii].

A **controlled vocabulary** is a way to describe knowledge for subsequent retrieval. It mandates the use of predefined, authorized terms that have been preselected by the designers in contrast to natural language vocabularies, which have no such restriction.

A **taxonomy** is a hierarchical classification of the terms in a controlled vocabulary. It captures no complex relationships between elements, except subclass/superclass relations.

An **ontology** is a formal naming and definition of the types, properties, and interrelationships of the entities that really or fundamentally exist for a particular domain[ix]. Ontologies aim to define which entities, provided with their associated semantics, are necessary for knowledge representation in a given context[x]. Ontologies and related information technology provide an opportunity to share a common understanding of the structure of information within a specific domain, the possibility to reuse domain knowledge, to make domain assumptions explicit and to analyse domain knowledge[xi].

4.10 References

i Review of Materials Modelling. What makes a material function. DOI: 10.2777/417118.
ii https://materialsmodelling.com/2016/12/20/cen-workshop-on-materials-modelling-terminology-classification-and-metadata/.
iii R. Arp, B. Smith, A.D. Spear. Building Ontologies with Basic Formal Ontology. MIT Press 2015. ISBN: 978-0-262-52781-1.
iv BFO Reference, see https://github.com/BFO-ontology/BFO
v ISO 15489-1 s 3.12
vi G.J. Schmitz et. al.: Towards a metadata scheme for the description of materials – the description of microstructures Science and Technology of Advanced Materials 17:1(2016) 410-430
vii http://www.dcc.ac.uk/resources/metadata-standards/cif-crystallographic-information-framework
viii Building a metadata schema – where to start. ISO/TC 46/SC11N800R1
ix https://en.wikipedia.org/wiki/Ontology_(information_science)
x Thomas R. Gruber (1993). Toward principles for the design of ontologies used for knowledge sharing. Originally in N. Guarino and R. Poli, (Eds.), International Workshop on Formal Ontology, Padova, Italy. Revised August 1993. Published in International Journal of Human-Computer Studies, Volume 43, Issue 5-6 Nov./Dec. 1995, Pages: 907-928, special issue on the role of formal ontology in the information technology.
xi David Lamas, Metadata and Ontologies, 2011; https://www.slideshare.net/davidlamas/metadata-and-ontologies

The outcome of this workshop was compiled by	Gerhard Goldbeck (Goldbeck Consulting Limited) Alexandra Simperler (Goldbeck Consulting Limited)

Lead beneficiary	GCL
Contributing beneficiaries	GCL, FRAUNHOFER

EC-Grant Agreement	723867
Project acronym	EMMC-CSA
Project title	European Materials Modelling Council - Network to capitalize on strong European position in materials modelling and to allow industry to reap the benefits
Instrument	CSA
Programme	HORIZON 2020
Client	European Commission
Start date of project	01 September 2016
Duration	36 months

Consortium		
TU WIEN	Technische Universität Wien	Austria
FRAUNHOFER	Fraunhofer Gesellschaft	Germany
GCL	Goldbeck Consulting Limited	United Kingdom
POLITO	Politecnico di Torino	Italy
UU	Uppsala Universitet	Sweden
DOW	Dow Benelux B.V.	Netherlands
EPFL	Ecole Polytechnique Federale de Lausanne	Switzerland
DPI	Dutch Polymer Institute	Netherlands
SINTEF	SINTEF AS	Norway
ACCESS e.V.	ACCESS e.V.	Germany
HZG	Helmholtz-Zentrum Geesthacht Zentrum für Material- und Küstenforschung GMBH	Germany
MDS	Materials Design S.A.R.L	France
Synopsys / QW	Synopsys (former QuantumWise A/S)	Denmark
GRANTA	Granta Design LTD	United Kingdom
UOY	University of York	United Kingdom

Coordinator – Administrative information	
Project coordinator name	Nadja ADAMOVIC
Project coordinator organization name	TU WIEN
Address	TU WIEN | E366 ISAS | Gusshausstr. 27-29 | 1040 Vienna | Austria
Phone Numbers	+43 (0)699-1-923-4300
Email	nadja.adamovic@tuwien.ac.at
Project web-sites & other access points	https://emmc.info/